2-P

NASA TECHNICAL TRANSLATION NASA TT F-14,846

EFFICIENCY STUDIES ON COCKPIT DISPLAYS

R. Beyer

Translation of "Wirksamkeitsuntersuchungen von Anzeigen für die Flugführung," fifth yearly meeting of the Deutsche Gesellschaft für Luft- und Raumfahrt e.V., Berlin, October 4-6, 1972, 30 pages

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D.C. 20546 APRIL 1973

32

Table of Contents

EFFICIENCY STUDIES ON COCKPIT DISPLAYS

Ralf Beyer

1. Introduction

The development of electronic display engineering has led to considerable expansion of the possibilities for representing information, already being used to a large extent, at least in military aviation. Electronic displays offer a number of advantages over conventional display methods which can be essentially characterized as follows:

a) Due to electronic image generation without moving parts, all problems and restrictions have been eliminated regarding the arrangement of many individual displays in a limited display field so as to be located close together and therefore easy to survey. Although there are no limitations placed on the combining of displays in a single display instrument, from the technical point of view, it is necessary to consider the risk that the comprehensibility of representation can suffer from excessive exploitation of this possibility.

b) Electronic displays make multiple utilization of the available display area possible, through the displaying and erasing of symbols, without additional space requirements on the instrument panel. For each phase of flight, for example, it is possible to erase all individual displays which are not needed and to show new displays, thereby increasing the overall comprehensibility of representation.

---

* Numbers in the margin indicate pagination in the foreign text.

c) The shape, size, color, orientation, position and surface structure of electronically generated symbols are subject to no restrictions. They can be varied dynamically and independently of one another. They permit an unlimited multiplicity of representation.

d) Electronic displays offer the possibility of integrating several pieces of information into a single display element. For example, a circular display character can be used to simultaneously indicate the position of the aircraft relative to the landing field (circle position, two parameters), altitude (circle diameter, one parameter) and time remaining to landing (specially labeled portion of the circle's circumference, one parameter).

e) With the aid of suitable instruments (light spots, controllable position indicators, contact wires), electronic displays can also be used for data input suitable for EDP.

Of these properties of electronic displays, the combination of readings in a spatially limited display field and the multiple utilization of available display areas are primarily being used for flight control at this time. The reason for limited utilization of the characteristics listed in paragraphs c)-e) is based on certain technical inadequacies of equipment which is presently available, on the sometimes underestimated efficiency of the available representational and informational input capabilities, and on a great shortage of acknowledged methods for comprehensively comparing the efficiencies of new modes of representation with displays already available.

The last-mentioned point will therefore be given special attention in the study described below. The possible applications of electronic displays, their technology, the methods of data coding, and the simulation of electronic displays for R&D purposes will be placed in the background. They are treated, along with other considerations, in [1].

2. The Evaluation of Displays

In an efficiency comparison between two displays, e.g. a conventional instrument with an electromechanical mode of operation and an electronic display, two criteria are of prime interest: 1. The effect which one display has on the accuracy of aircraft control in comparison to the other and 2. what changes in the pilot's workload result when it is used. A number of boundary conditions can be derived from this for the remarks to be discussed below:

1. The studies considered here involve relative quantities only, i.e., a display is evaluated in relation to the display used for reference purposes.

2. The accuracy of aircraft control and the pilot's workload possess a mutual interdependence which has so far been clarified only to an imperfect degree and which is affected only partly by the quality of the display.

3. The term "control accuracy" should not be taken in too restricted a sense, but includes not only the achievement of certain flight parameters but also variability in the implementation of flight maneuvers without specific reference quantities.

4. In the estimation of pilot workload, the inter-individual differences among test subjects which might occur here, or even their basic suitability, are of less interest than the reaction of test subjects to different displays. The first-named criteria

are employed prior to the studies, however, for collecting a group of subjects which is as representative and/or homogeneous as possible.

The goal of comparing the efficiencies of displays is to try to find that display with which an acceptable level of control accuracy can be achieved and, simultaneously, with which the work which the pilot must perform to attain this can be limited to the level which is absolutely necessary. It appears logical here, at the outset, to carry out the experimental studies somewhere where all factors which might affect the study are represented to a sufficient extent. Accordingly, simple legibility studies would be initially carried out on displays in a stationary test stand (fixed-cockpit flight simulator), and studies more closely related to flight control would take place in a moving-cockpit flight simulator, in which, among other things, experienced pilots are capable of properly utilizing the impressions of movement, as described in [2]. Finally, a test aircraft would be selected for studies under very realistic flight conditions. It has been found, however, that such a choice of test vehicle does not always result in success, and that a test carried out in a flight simulator under highly reproducible ambient conditions occasionally produces better results than a test flight.

Only those datas are treated below which are used in efficiency studies on displays. They are the following:

a) Control-technology data
b) Experimental-psychology data
c) Physiological data
d) Information obtained from pilots

The examples given below may be considered representative of the large number of types of data which cannot be mentioned here, due purely to reasons of scope.

2.1. Control-Technology Data

This refers to those data which are obtained by measurements taken on flight equipment, the navigational system, control units, or the like. Among other things, they include the maintaining of certain flight parameters such as position, heading, speed, the time required for a certain flight manuever, and so on. The selection of the parameter to be considered in a display comparison depends upon the particular flight manuever and, of course, upon the criteria applied to it. For fixed reference parameters, e.g. the ILS glide path, RMS values or absolute average error can be suitable measures for evaluating the experimental data obtained here, as well as those described in the following sections. Even without specific reference quantities, however, the magnitude of standard deviation can provide important information on the variability of the manuever being studied for different display instruments. But even when a criterion and the parameter to be measured are selected carefully, data of the type mentioned above often possess two little sensitivity for display studies to allow a comprehensive quality comparison between two displays. This should not mean, however, that simple control-technology data such as average error (AE) are not very effective in each case. Thus, for example, a constant relative deviation from the flight path during ILS approach caused by parallax errors was observed in [3] with a conventional display instrument, while no such errors occurred with the electronic display instrument used in the comparison flights. Moreover, the localizer and glide scope readings of the two instruments studied differed because of the greater precision of the electronic display, since there are no restrictions on indicator-needle movements here, and a greater maximum indicator deflection can be realized here for a given indicating surface and a given indicating range.

Inadequacies in the readings which may occur are often compensated for in their effect upon the accuracy of aircraft control by increased efforts on the part of the pilot, so that no significant differences in information are obtained from control-technology data. Test flights with original Pembroke instrumentation and with partial replacement of the instruments with an appropriate electronic display [3] have confirmed this state of affairs. Experience gained to date indicates that three things play a role here:

a) Efficiency studies on displays in the flight simulator [3] have shown that the above-named control-technology variables apparently have only a very low degree of sensitivity with regard to differences in display design. Although changes in measured values resulting from display changes could be demonstrated, they were all statistically insignificant, even under highly reproducible test conditions. Errors of "gross negligence" in the designing of displays will definitely cause significant differences in measured data, but they were not the subject of this study.

b) In flight tests [3], too, data differences of the type mentioned are hardly to be expected in such studies. Even if -- vis-a-vis tests in the flight simulator -- slight differences in measured data might be obtained between two displays being studied, due to a certain risk stress and the realistic impressions of motion and sound, their statistical significance is very probably nullified again by the increased variability of test-flight conditions.

c) Flight manuevers carried out during studies to date, such as maintenance of altitude or heading, ILS approach, and the like, probably do not represent such difficult tasks for the experienced pilot that he must absolutely rely on displays of anthropotechnically optimum design and must suffer the consequences of significantly poorer control data for readings which are of a qualitatively inferior level. Flight tests carried out under realistically simulated IFR conditions [3] with orange/blue filters in the cockpit glass or pilots' glasses gave support, on the basis of a large number of details, to the assumption that during ILS approach, the pilots were still quite far from the workload limit at which maximum demands are made on the proper human engineering of display equipment and at which changes in display would thus have to directly affect control precision.

Viewed overall, the pilot's average workload appears to be decisive for the demonstrability of control errors caused by displays in such studies and in a simple analysis of control-technology data. Whether it can be raised to a usable value for study purposes of the type described above is doubtful just on the basis of safety-engineering considerations.

On the one hand, a relatively simple evaluation of control-technology data thus does not appear to be all too effective for comparing displays; on the other, an effect on these values caused by the display cannot be ruled out. This effect is now being studied with more extensive processing of the measured data. Thus a study has been made in [4], for example, as to the intercorrelations which exist between the effective values of error in the angles of yaw and pitch when different displays are used. When the standard instruments of a C-131 aircraft were used here, it was found that a small but statistically significant positive intercorrelation existed between the two above-mentioned parameters. This means that an increase in angular error about one axis is related, from a statistical point of view, to an increase in angular error about the other. When a display instrument of similar appearance (Kaiser FP 50) was used, a similar correlation could be demonstrated, but of negative sign. An increase in the angular error about one axis is thus, statistically speaking, related to a reduction in the angular error about the other axis here. Similar results were obtained with other parameters. Such analyses thus appear suitable for obtaining a certain insight into the pilot's operating strategy connected with the particular display instrument. Although an effect by the display on aircraft control could thus be proven with the aid of special analysis techniques in the example given, this proof would not have been possible if the statistically insignificant differences in effective values of the pertinent angular errors had merely been considered. In [5], an attempt was made, through a correlation analysis of control and adjustment signals, to determine the quantity of data "processed" by the pilot per unit time. Here, too, an effect resulting from the display, in this case a combination of several displays in a common display field which could be easily surveyed, could be well demonstrated. Again, such proof would definitely not have been possible in a less extensive analysis of control and adjustment signals.[1]

* * *

Experiments with primary and secondary tasks are based on certain assumptions which can limit the applicability of the results obtained. The most important are the following:

-- The primary task always has a higher priority than the secondary task. The pilot thus concerns himself with the secondary task only when the primary task permits. The maintenance of this prerequisite during testing can hardly be exactly demonstrated, however.

---

[1] [Portion missing from German document.]

-- The primary and secondary tasks, combined, result in 100% utilization of the pilot's performance. Poor performance with regard to the secondary task is caused only by a higher level of difficulty of the primary task or by an unsuitable display design. However, this assumption depends upon the differences in test-subject fatigue and motivation with respect to each subtask.

-- The primary and secondary tasks do not affect one another. The strategy applied to solve the primary task and the performance which is executed here can be considerably influenced by the type of secondary task if the experiment is poorly laid out.

An attempt can be made to largely satisfy one of the above assumptions by specially designing the experiment. In order to actually achieve a secondary task which absorbs the reserve capacity of the test subject, it is possible to control its level of difficulty as a function of the performance executed in connection with the primary task (cross-adaptive secondary task [7, 8]). When the display instrument is changed here, for example, the secondary task's level of difficulty is automatically adapted to the new conditions, so that the test subject has an approximately constant workload. The magnitude of the workload itself is determined by a predetermined performance level for the primary task, upon the exceeding of which the secondary task becomes more difficult, and vice versa. Such experiments are likewise not free of limitations, of course, e.g. their low compatability with practical flight tasks. Nevertheless, the expectations placed on them are high, particularly in terms of a sensitive method for determining relative workload in display studies which guarantees constant working conditions.

Another possibility for comparing the efficiency of displays is measurement of the test subject's reaction speeds to displays with different designs. Thus in experiments with monochromatic and polychromatic displays, for example, it was found that under certain experimental condtions, brightness-coded displays resulted in a longer reaction time on the part of test subjects than color-coded displays, even though the number of perceptive errors was not different for the two types of coding [6]. One of the possible reasons for this may have been the fact that a given brightness can hardly be correctly identified without the presence of a reference brightness level, whereas colors can normally be correctly recognized without a color reference.

Another type of experiment uses only brief presentation of the display (tachistoscopy), ranging from about 10 ms to several seconds. The aim of the experiment is to test recognition of the display or individual subelements under impaired visibility conditions. Such experiments are described in [9] in which the exposure times were progressively shortened. One of the two displays studied exhibited a stronger dependence, with regard to erroneous interpretations, upon exposure time than the other, i.e., it became unusable sooner as exposure time was decreased. In [10], tachistoscopy is used to perform a comparison between a conventional instrument panel of a Pembroke and the same instrument panel with the most important instruments replaced with an electronic display (Fig. 1). In contrast to the previously mentioned experiment, however, exposure time was kept constant here at several seconds, and the number of correctly recognized pieces of information was used as the criterion. It was found that when the conventional instruments were partially replaced with a corresponding electronic display, approximately 36% better identification performance was achieved. In a similar comparison between the electronic display and a modern Attitude Director

Indicator of conventional design (Fig. 2), about 40% better identification performance was recorded for the electronic display than for the conventional indicating instrument. These results contrast with analysis of the measured control-technology data described above, in which average error and variance yielded no significant differences in measured values with regard to the various forms of display.

As is true of the experiments with primary and secondary tasks, tachistoscopic experiments are likewise not free of uncertainties, but these are not so extensive, in numerical terms, and appear to be more comprehensible than those in experiments with primary and secondary tasks. They include, among other things, the relatively unrealistic -- as compared to practical flight conditions -- operating situations in which the test subjects find themselves during the experiment and the fact that the results obtained under limiting conditions cannot be immediately and directly applied to the normal case: For example, if the exposure time for which a display is just barely readable is taken as the criterion, this by no means says that a display which is no longer readable for this exposure time is also to be interpreted as worse for continuous representation than the display employed as a reference. Moreover, when continuous representation is used, for example, it is also possible to perceive not only the position of a display element but also its movement. During flight, additional information is also obtained from the movements of the aircraft, whereby probability increases for a given display-element position when viewed the next time, and proper reading of the instrument is promoted relative to the tachistoscopic experiment.

Although the experimental-psychology data thus still involves many effects which can only be coped with poorly, if at all, such data are being given increasing importance as indirect indicators of the pilot workload in display studies. In spite of many advances aimed at the direct measurement of psychological stress (see [11], among others), the demand for such simple approaches, even though involving only indirect measurement, will remain high until acknowledged and feasible methods become available.

2.3. Physiological Measurements

Another method of estimating psychological stress on the pilot for the purpose of comparing the efficiencies of displays is based on psychophysiological phenomena. For example, it can be demonstrated that the electrical resistance of the skin, the blink rate, the potentials which can be measured on the body, variations in pulse rate, etc., are situation-dependent (fright, joy, work involving concentration, and so on). A study of the articles available in this field suggests that many of the parameters which are atuall feasible have already been investigated at some time in some form, and in many cases possess a statistically significant correlation with regard to variation of the test subject' workload. On the other hand, many studies, such as [12] indicate that certain variables remain proportional to one another over a large measurement range, and thus a large number of the variables can be dispensed with, under some circumstances, in favor of a few or even one. In flight tests, moreover, it has been possible to demonstrate the quite high sensitivity of many variables vis-a-vis ambient conditions. For example, measurement of the psychogalvanic reflex in a display comparison during landing approach yielded fluctuations in the measured values which were one to two orders of magnitude greater than was expected for a change in display instruments. However, these variations were caused primarily by the changing conditions under which the flight tests were made (gusts, observation of other traffic, etc.). Differences in measurement values which might have actually existed and which could have been attributed to a change in displays were thereby completely masked. Previous experience indicates that such a large number of known and unknown influencing factors are decisive in physiological measurements that display-related affects probably assume only a subordinate position. It follows from this that experimental conditions which can be monitored with particular accuracy should be present for physiological measurements. But this requirement is met to a much higher degree on the ground, e.g. in a flight simulator, than in the flight test. If physiological data are used for a display efficiency comparison, statistically significant results will still be obtained in laboratory tests for some time to come, even if not very realistic. This by no means indicates that flight tests would not be necessary. But they serve more as a validation of psychophysiological measurement procedures for suitability studies and the like than studies of the efficiency of displays. Thus while it has been possible to adequately demonstrate their stability and suitability many times in experimental-psychology tests, we are still at the beginning of development concerning the physiological variables in this regard.

2.4. Subjective Evaluations

The pilot is in the unique position of being able to give an overall evaluation of the displays studied and thus to participate in decision-making with regard to preference for one display over another. While the study is being carried out, he can recognize influencing factors which were overlooked at the beginning of the experiment and which might only have been discovered with the aid of an analysis of the variables described above. Many of these effects would probably always go unnoticed without an evaluation of the display by the pilot, however. On the other hand, the judgment of the pilot is determined by a series of very different factors which come to bear during a display efficiency comparison in flight tests. Among other things, these include risk stress, the work done in executing the flight mission, the technical perfection and appearance of the display instrument, a certain positive or negative bias for or against the display instrument, transference of the experience obtained with the studied display during flight tests to other practical situations not studied there, and so on. An attempt must now be made, using suitable tests, to determine the effects which these factors have upon overall evaluation. So-called open-ended questionnaires permit extensive characterization of all factors which influence display evaluation. But the answers are as a rule difficult to quantify and thus yield only conditionally comparable results. Questions with possible answers prespecified do not have this disadvantage, but the danger exists that important factors will not be addressed at all in the applicable tests. Nevertheless, the last-mentioned form of test is usually used in display studies.

In the pair-comparison method, the test subject is shown two displays in each case, of which one is to be given preference. In a study on the effect of the coloring of displays [6], for example, it was determined that a group of 22 engineers and private pilots and a group of 25 professional pilots preferred the display in Fig. 3 a to the display in Fig. 3 b, 18 to 4 and 21 to 4, respectively. The corresponding numbers for the displays in Figs. 3 c/d are 19 to 3 and 20 to 5, respectively, in favor of a display like that in Fig. 3 c. On the other hand, different evaluations are also obtained which are apparently related to the educational background and activities of the test subjects. In a comparison of the displays shown in Figs. 3 a and 3 c by the engineers and private pilots, for example, the display shown in Fig. 3 a was preferred 18 to 4, whereas the professional pilots gave a more balanced evaluation of 13 to 12.

Another possible method for determining the consensus regarding the displays studied is to employ the "semantic differential." A number of paired adjectives of opposite meaning are used here which can describe the characteristics of the display to be studied and which are separated by a graduated scale. Examples of such adjectives might be: accurate - inaccurate, fast indicating - sluggish, etc. On the basis of his evaluation of the display, the test subject must then mark an appropriate point on the graduated scale for each pair of adjectives. The position of this point depends upon whether the subject prefers one or the other of the adjectives to describe his impression. All of the answers taken together thus yield an impression of the overall evaluation of the display studied; serious design errors or particularly appreciated characteristics of the display can simultaneously be determined relatively clearly. Difficulties in setting up such tests are presented by the search for adjectives in each pair which are truly opposites, and the coverage of all properties to be studied by means of an adequate number of relevant characterizing terms.

A study method which is similar to the test described above utilizes a so-called "polarity profile" to determine positive and negative characteristics of the display. Each of the details to be studied here is assigned an evaluation scale, e.g. from -5 to +5 (Table 3). The pilot then checks off that position on the scale which best represents his impression regarding the display being studied and regarding the particular detail, as compared with a display used as the reference. Before the test is begun, the pilots are instructed, among other things, to use an evaluation of "0" only in exceptional cases. The results shown in Table 3 come from a comparative study of the displays shown in Fig. 2. The results indicate that the electronic display can elicit a larger number of positive, and more highly positive, evaluations than the conventional display instrument.

Three factors appreciably affect the unbiased evaluation of displays by pilots:

-- Inadequate knowledge and positive or negative pilot prejudice regarding the display to be evaluated, which can change with increasing experience.

-- The pilot's desire not to disappoint the developer of displays in his efforts to achieve better configurations.

-- A certain tendency to blame problems which arise during flight with the display being evaluated on the display rather than on himself.

A special questionnaire technique makes use of the last two points described above by asking the question of interest twice. Such a pair of questions might read like this, for example: Where the displayed characters large enough? Did you ever have problems using the display which could be attributed to inadequate character size? Both questions are presented in hidden form within the framework of a questionnaire; with four possible reply combinations, answering them can yield replies which are relatively poorly established in two cases and replies which are relatively well established in two cases. Table 1 shows the results obtained in an efficiency comparison between an electronic display (Fig. 1) and the corresponding conventional display instruments of a Pembroke. The table gives the number of the particular question in the first column, the number of the corresponding counter-question in the second, the question in abbreviated form in the third, and the positive or negative evaluation given by each pilot in the following columns, including cases which could not be decided by the pilot. Table 2 shows the results of the evaluation, in which the positive or negative evaluations obtained from the replies to a pair of questions are indicated by a plus or minus sign, while a zero represents a contradictory reply to the pair of questions. If one of a pair of questions was answered with "do not know," both questions or the replies given were indicated, with a question mark, as being unsuitable for evaluation. The positive and negative assessments determined for each complex of questions were then added algebraically and entered as the evaluation in the far right column of the table. The results essentially indicate three things: The inappropriateness of the angle of roll display element used in the electronic display, the apparently adequate to good quality of the other display elements and characteristics of the electronic display, and the ineffectiveness of a comparison, by the pilots, of the workload connected with the use of the electronic display and of the corresponding conventional display instrument. On the one hand, the last-mentioned observation would depend to a great extent upon the design of the questionnaire; on the other, a comparison of workloads by the individuals involved appears to be just as problematic as the estimation of psychological stress is anyway.

3. Closing Remarks

The few examples of the variables used in display efficiency studies should have made clear how complex the evaluation of displays can be and which tasks are urgent for the further development of evaluation methods. Since a large number of parameters and measurement methods have been developed in the past, further enlargement of this number should take a subordinate role in favor of improvements in stability, reliability, sensitivity, and relevance of the test methods. Many research systems already meet this requirement in that we are limiting ourselves to just a few variables for a given outlay, but are analyzing these that much more thoroughly and testing them under the widest variety of conditions. Well-planned division of labor between the participating institutions would certainly be a great help here in advancing the level of knowledge in this field over a wide front. In spite of the limitation on means, a certain equilibrium should be maintained here between control-technology, experimental-psychology, physiological and subjective parameters as long as we are not sure that a certain area cannot provide an essential contribution to the evaluation of displays. In the absence, even today, of a relative weight distribution with respect to the various parameters within the framework of an overall evaluation, experience obtained to date indicates the importance of supporting evaluations with several different parameters. The goal of the study here is to have different variables available which each support the acceptance of display "A" over display "B." In practice, however, the results often have a much less unequivocal character. For example, it is not unusual for the control-technology variables to indicate no difference between "A" and "B," an experimental-psychology test to support "B," physiological variables to be incapable of evaluation, and the subjective evaluation of the displays by pilots to support the acceptance of "A." The only things which help here are to uncover as yet unknown effects, eliminate factors which cannot be monitored and the variabilities which they cause, select test subjects in a better way, employ more relevant and sensitive evaluation methods, etc. It thereby often proves possible to bring out differences in efficiency between two different displays which had perhaps been suspected but had not yet been quite demonstrable. It may then be that several parameters will indicate the approval of display "A" over "B," so that a decision between "A" and "B" which is well established in a manner commensurate with the state of the art becomes possible. Of course, this approach sometimes results in a considerable outlay, the result of which is that we occasionally allow ourselves to be satisfied with a display evaluation which has a less solid basis. As long as no far-reaching conclusions for future developments are drawn from this,

this might be understandable on the basis of economy and calculated risk. Reservations must be expressed, however, if such display evalutions are made the foundation of new developments without testing to determine under what conditions the display used as a reference had itself been accepted.

Display evalutions usually have a measure of the statistical significance of the results obtained which, however, can by no means be considered equivalent to operational significance. Aside from the fact that the results of laboratory studies are not referred to practice in all cases, it is of interest to determine, for example, the gain provided by a preference for a polychromatic conventional display over a monochromatic electronic display which, while statistically significant, only took the impression of color into consideration, and not the above-mentioned advantages of the electronic display. Another example: according to [14], shape-coded test characters result in a reaction time which is longer by almost a factor of three than color-coded test characters; according to [15], however, they can be used to achieve a 30% higher number of usable coding possibilities. Moreover, shape coding is considerably easier to realize in electronic head-up displays, for example, than color coding. So which provides the greater overall gain here? Precise knowledge of the particular application and the technical possibilities for realization is therefore of considerable advantage in display efficiency studies, along with basic knowledge, to permit the development of displays to also be supported with practical information. In addition, suitable systems such as display simulators are necessary if the desired goal is also to be achieved while observing the economy aspect [13]. To be sure, even the most expensive systems would be useless if suitable programs were not available for their operation. A plan was therefore worked out more than 6 years ago within the framework of these studies, in which engineers and psychologists consider the tasks presented here as a unit and attempt to solve them in combination, using the most up-to-date, extensively software-oriented aids, in connection with the diverse facilities of a research laboratory for aeronautics and astronautics. This means that each worker remains primarily associated with the task, from formulation of the problem to its solution, and a team is not subdivided, for example, into groups assigned to theory, experiment, programming, evaluation, psychology, information theory, etc. Granted, a high level of universal and interdisciplinary performance is required from individual workers but is particularly important precisely in the field being discussed here and has its appeal in continual change, from flight testing to laboratory investigation, from computer-supported instrument and display design to psychological testing, from the long-term research program to quick consultation with a customer. The correctness of this plan has been confirmed by the success which has been achieved in the past.

4. Summary

As electronic display techniques are being developed and introduced, the need for information on the efficiency of the new display methods, particularly in comparison with techniques used previously, is also increasing. At the same time, the development of suitable evaluation methods for display studies has experienced a strong upswing. Since a large number of extremely varied evaluation parameters are already known and being applied, attempts are now being made to prove their stability, reliability, sensitivity and relevance. A number of evaluation parameters have been presented, as representative of many others, the results obtained with them in experimental studies have been used for the purpose of illustration, and a number of the as yet unsolved problems have been discussed in each case. This report is meant to contribute to our understanding of the

capabilities and of the significance of display efficiency studies, but at the same time also to stimulate improvement of the techniques presently available. The saying, "Something better is the enemy of something good," is particularly applicable to this field, still undergoing intensive development; not just the methods and their implementation but also the knowledge and capabilities of the scientists who are active in this field are being called into play.

REFERENCES

1. Beyer, R., "Electronically generated data representation in aircraft," DFVLR, DLR FB 72-43, 1972.

2. Dierke, R. and Erdmann, F., "The effect of impressions of motion upon pilot error in simulated ILS approaches," DFVLR, DLR Mitt. 70-29, May 1971.

3. Kohnen, E. and Schenk, H.-D., "Comparison of conventional and electronic display instruments in the flight simulator and test aircraft," DFVLR, Institut für Flugführung, March 1972.

4. Cross, K. D. and Bittner, A. C., "In-flight comparison of the Kaiser FP-50 flight director with standard C-131 instruments," JANAIR Report No. 690412, March 1970.

5. Endrass, H., Etschberger, K., and Maier, P., Kybernetische Grundsatzuntersuchungen zur Auslegung von Instrumentierungs- und Regelungs-systemen, [Basic Cybernetic Studies on the Design of Instrumentation and Control Systems], Dornier, March 1970.

6. Beyer, R., Schenk, H.-D., and Zietlow, E., "Readability and interpretation studies on electronic displays: efficiency studies on the brightness coding and color coding of display elements," DFVLR, DLR FB 71-57, 1971.

7. Kelley, C. R. and Wargo, M. J., "Cross-adaptive operator loading tasks," Human factors 5, No. 9, 395-404 (1967).

8. Clement, W. F., Allen, R. W., and Graham, D., "Pilot experiments for a theory of integrated display format," System Technology Inc., JANAIR Report 711107, October 1971.

9. Berström, B., "Interpretability studies of electronic flight instruments," SAAB Aktiebolag, TN 61, 1967.

10. Heinze, W., "Tachistoscopic studies on electronic and conventional cockpit displays," DFVLR, Institut für Flugführung, 1972.

11. Spyker, D. A., Stackhouse, S. P., Khalafalla, A. S., and McLane, R. C., "Development of techniques for measuring pilot workload," NASA Contractor Report CR-1888, November 1971.

12. Morehouse, L. E., "Model of a system utilizing heart rate to monitor man at work in an alien environment," IEEE Conference Record No. 69C58-MMS, 1969.

13. Beyer, R., "Display simulation and display flight testing techniques," Zeitschrift für Flugwissenschaften 18, No. 11 (1970).

14. Williams, L. G., "The effects of target specifications on objects fixated during visual search," Attention and Performance I, North Holland Publishing Company, Amsterdam, 1970.

15. Morgan, C. T., Human Engineering Guide to Equipment Design, McGraw Hill, 1963.

6. Figures and Tables

Fig. 1. Electronic display on the instrument panel of a Pembroke.

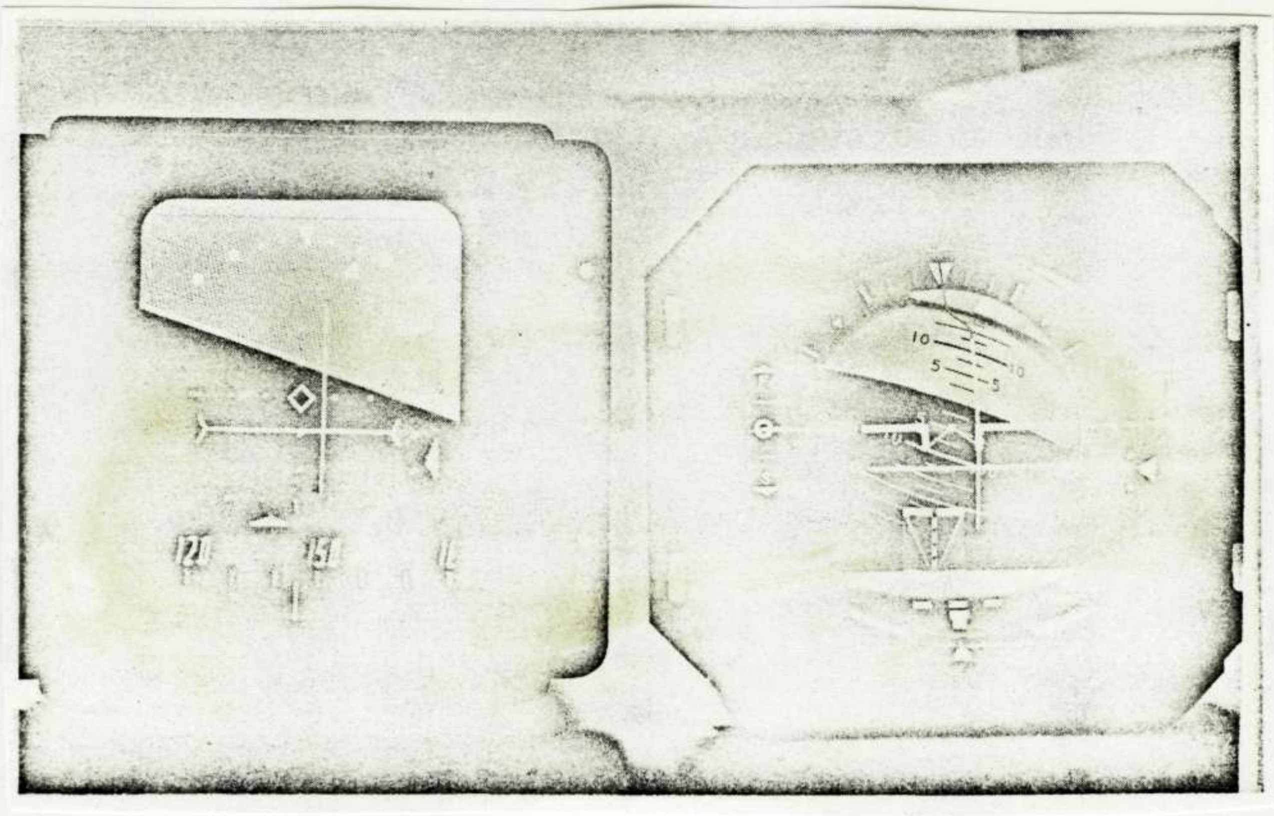

Fig. 2. Modern display instrument of conventional design (right, Lear Siegler 4058 AC) and electronic display instrument (left, DFVLR).

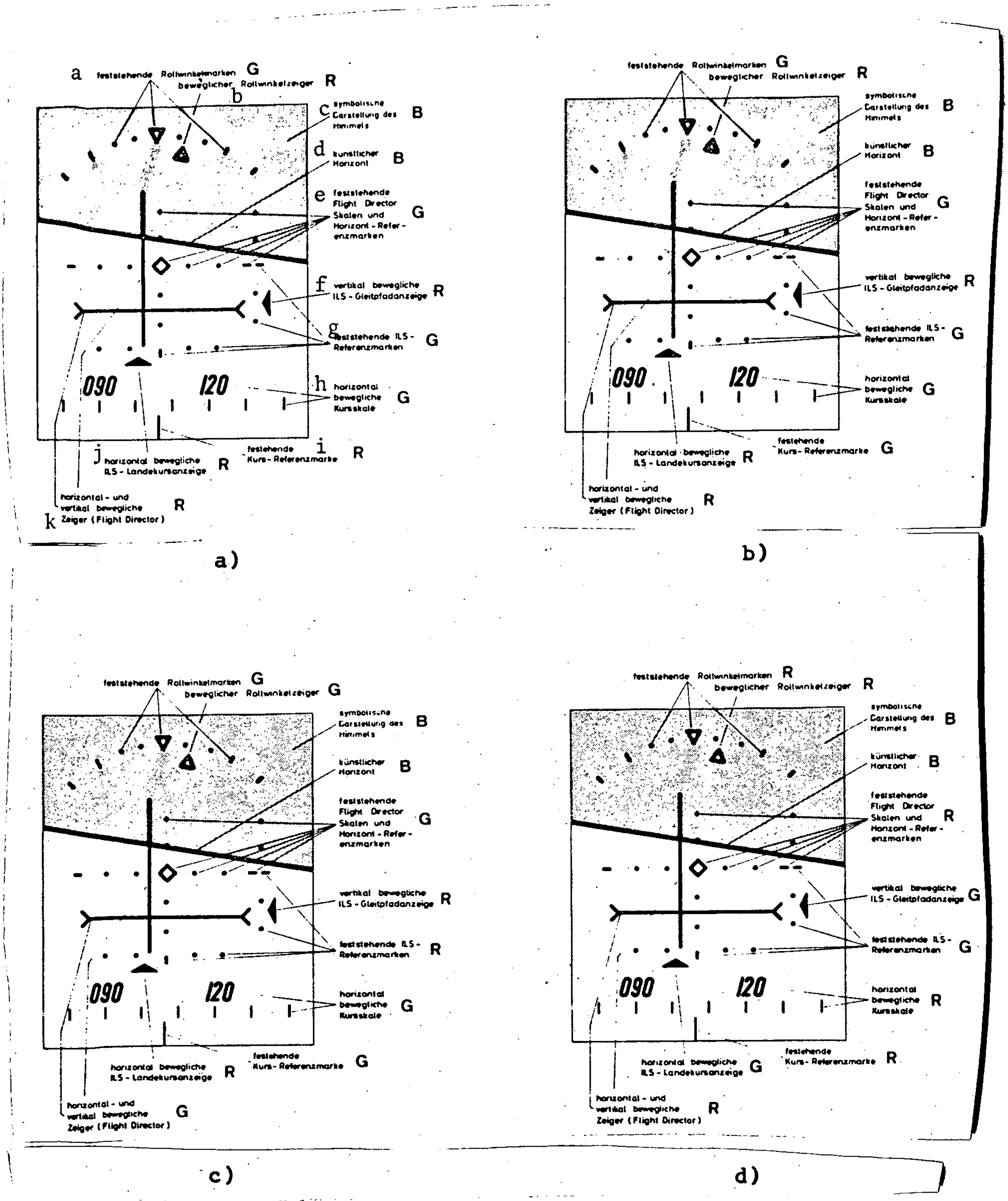

Fig. 3. Different coloring schemes for an electronic display (see Section 2.4). R = red, G = green, B = blue.

[Key on following page]

Key to Fig. 3.

a. Fixed angle of roll marks
b. Moving angle of roll indicator
c. Symbolic representation of sky
d. Artificial horizon
e. Fixed Flight Director scales and horizon reference marks
f. Vertically moving ILS glide path indicator
g. Fixed ILS reference marks
h. Horizontally moving heading scale
i. Fixed heading reference mark
j. Horizontally moving ILS landing path indicator
k. Horizontally and vertically moving indicators (Flight Director)

TABLE 1. RESULTS FROM A QUESTIONNAIRE TEST (SEE SECTION 2.4)

Question	Assoc'd question	Question in abbreviated form	Reply											
			Yes				?				No			
			Pilot 1	Pilot 2	Pilot 3	Pilot 4	Pilot 1	Pilot 2	Pilot 3	Pilot 4	Pilot 1	Pilot 2	Pilot 3	Pilot 4
1	22	Display airworthy?	x	x	x	x								
2	16	Image brightness adequate?	x	x	x	x								
3	17	Heading indicator easily readable?	x	x	x	x								
4	21	Better approaches possible with roll angle ind'r?	x		x	x						x		
5	20	Theoretical course setting OK?	x	x	x	x								
6	12	Another display arrangement more airworthy?				x					x	x	x	
7	14	Pilot error caused by type of pitch angle scale?					x					x	x	x
8	26	Image free of flicker & vibration?	x	x		x							x	
9	28	Intercept angle display OK?	x		x	x						x		
10	18	Keep std. instruments as back-up?	x			x						x	x	
11	25	Workload noticeably reduced by electronic display?	x			x						x	x	
12	6	Arrangement of electronic display on instrument panel OK?		x	x	x					x			
13	24	Problems reading heading?									x	x	x	x
14	7	Pitch angle easily readable?	x	x		x			x					
15	27	Better approaches possible after longer training?	x	x	x					x				
16	2	Image brightness often inadequate?									x	x	x	x
17	3	Should heading ind'r be more precise?	x									x	x	x
18	10	Confident in electronic display?	x	x	x	x								
19	23	Horizon line bright enough?	x	x	x	x								
20	5	Must improve theoretical course setting?					x					x	x	x
21	4	Roll angle easily readable?				x					x	x	x	
22	1	Any problems with the electronic display?			x						x	x		x
23	19	Must horizon line be shown clearer?			x						x	x		x
24	13	Heading easily readable?	x	x	x	x								
25	11	Does electronic display impair attention to other displays?	x		x							x		x
26	8	Disturbed by image flickering?									x	x	x	x
27	15	Approach with electronic display just as good as with standard instrument?	x	x	x	x								
28	9	Problems in determining intercept angle?									x	x	x	x

TABLE 2. EVALUATION OF DATA IN TABLE 1

Reference No.,	Assoc'd questions	Complex of questions	Pilot 1	Pilot 2	Pilot 3	Pilot 4	Evaluation
1	02/16	Image brightness	+	+	+	+	+4
2	24/13	Heading indicator	+	+	+	+	+4
3	01/22	Airworthiness	+	+	0	+	+3
4	03/17	Heading indicator	0	+	+	+	+3
5	27/15	Comparison of airworthiness of electronic display and of standard instruments	+	+	+	?	+3
6	08/26	Image stability	+	+	0	+	+3
7	19/23	Artificial horizon	+	+	0	+	+3
8	05/20	Theor. course setting	?	+	+	+	+3
9	09/28	Intercept angle determination	+	0	+	+	+3
10	14/07	Pitch angle reading	?	+	?	+	+2
11	18/10	Confidence in electr. display	0	+	+	0	+2
12	12/6	Config. on instrument panel	0	+	+	0	+2
13	11/25	Workload	0	-	0	+	0
14	21/04	Roll angle reading	-	0	-	0	-2

TABLE 3. POLARITY PROFILE (SEE SECTION 2.4). EVALUATION OF TWO DISPLAYS BY THREE PILOTS

	Conventional display											Electronic display										
	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Gen. symbol design				x					x		x					x				x	x	
Gen. reading accuracy			x	x					x								x		x	x		
Reading accuracy with indicator at zero					x	x			x							x			x	x		
Roll angle reading							x	x	x						x	x				x		
Pitch angle reading				x				x	x								x		x x			
Localizer reading			x				x	x								x			x			x
Glide slope reading			x		x		x									x			x			x
Flight director reading				x	x				x							x					x	x
Image contrast/illum.						x					x x		x						x	x		
∑	0	0	3	4	3	2	3	3	6	0	3	0	1	0	1	6	2	0	7	5	2	3
∑			10			2			15					8			2				17	